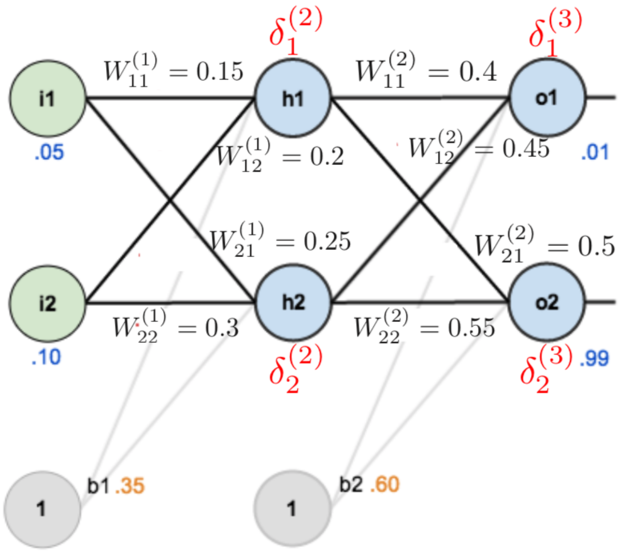
Practice Problem

Perform one iteration of forward propagation and backpropagation in the following neural network, assuming learning rate $\alpha = 0.5$. We also assume a sigmoid activation function $g(x)$ for the hidden nodes, where $g'(x) = g(x)[1 - g(x)]$.



Backpropagation Implementation

- For each node i in output layer L: $\delta_i^{(L)} = (\alpha_i^{(L)} - y_n) f'(z_i^{(L)})$
- For each (hidden) node i in layer $l = L-1, L-2, \ldots, 2$: $\delta_i^{(l)} = \left(\sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)} \right) f'(z_i^{(l)})$
- Compute the desired partial derivatives as: $\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{ij}^{(l)}} = \alpha_j^{(l)} \delta_i^{(l+1)}$, $\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta b_i^{(l)}} = \delta_i^{(l+1)}$
- Update the weights as: $W_{ij}^{(l)} := W_{ij}^{(l)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{ij}^{(l)}}$, $b_i^{(l)} := b_i^{(l)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta b_i^{(l)}}$

$$g(x) = \frac{1}{1+e^{-x}} \quad , \quad g'(x) = g(x)\left[1-g(x)\right]$$

Forward propagation

$$h_1 = g\Big(\underbrace{W_{11}^{(1)} i_1 + W_{12}^{(1)} i_2 + b_1}_{z_1^{(2)}}\Big) = g(0.15\times0.05 + 0.2\times0.10 + 0.35) = 0.5932699$$

$$h_2 = g\Big(\underbrace{W_{21}^{(1)} i_1 + W_{22}^{(1)} i_2 + b_1}_{z_2^{(2)}}\Big) = g(0.25\times0.05 + 0.3\times0.10 + 0.35) = 0.596884$$

$$o_1 = g\Big(\underbrace{W_{11}^{(2)} h_1 + W_{12}^{(2)} h_2 + b_2}_{z_1^{(3)}}\Big) = g(0.4\times0.5932 + 0.45\times0.5968 + 0.60) = 0.7513$$

$$o_2 = g\Big(\underbrace{W_{21}^{(2)} h_1 + W_{22}^{(2)} h_2 + b_2}_{z_2^{(3)}}\Big) = g(0.5\times0.5932 + 0.55\times0.5968 + 0.60) = 0.7729$$

Backpropagation

$$\delta_1^{(3)} = (o_1 - y_1)\, g'(z_1^{(3)}) = (0.7513 - 0.01)\, g'(0.4\times0.5932 + 0.45\times0.5968 + 0.60) = 0.1384$$

$$\frac{\partial J(W,\vec{b})}{\partial W_{11}^{(2)}} = h_1 \delta_1^{(3)} = 0.082167$$

$$W_{11}^{(2)} := W_{11}^{(2)} - \alpha \frac{\partial J(W,\vec{b})}{\partial W_{11}^{(2)}} = 0.4 - 0.5\times0.082167 = 0.3589$$

$$\frac{\partial J(W,\vec{b})}{\partial W_{12}^{(2)}} = h_2 \delta_1^{(3)} = 0.082667$$

$$W_{12}^{(2)} := W_{12}^{(2)} - \alpha \frac{\partial J(W,\vec{b})}{\partial W_{12}^{(2)}} = 0.45 \times 0.5 \times 0.082667 = 0.4086$$

$$\delta_2^{(3)} = (o_2 - y_2)\, g'(z_2^{(3)}) = (0.7729 - 0.99)\, g'(0.5\times0.5932 + 0.55\times0.5968 + 0.60) = -0.038098$$

$$\frac{\partial J(W,\vec{b})}{\partial W_{21}^{(2)}} = h_1 \delta_2^{(3)} = -0.02260$$

$$W_{21}^{(2)} := W_{21}^{(2)} - \alpha \frac{\partial J(W,\vec{b})}{\partial W_{21}^{(2)}} = 0.5 - 0.5(-0.022) = 0.51130$$

$$\frac{\partial J(W,\vec{b})}{\partial W_{22}^{(2)}} = h_2 \delta_2^{(3)} = -0.02244$$

$$W_{22}^{(2)} := W_{22}^{(2)} - \alpha \frac{\partial J(W,\vec{b})}{\partial W_{22}^{(2)}} = 0.55 - 0.5(-0.02244) = 0.561370$$

$$\delta_1^{(2)} = \left(w_{11}^{(2)} \, \delta_1^{(3)} + W_{21}^{(2)} \delta_2^{(3)} \right) g' \left(z_1^{(2)} \right)$$

$$= \left[0.4 \times 0.138498 + 0.5 \times (-0.038098) \right] g' \left(0.15 \times 0.05 + 0.2 \times 0.1 + 0.35 \right)$$

$$= 0.008771$$

$$\frac{\partial J(w, \vec{b})}{\partial W_{11}^{(1)}} = i_1 \cdot \delta_1^{(2)} = 0.05 \times 0.008771 = 0.0004385$$

$$w_{11}^{(1)} := w_{11}^{(1)} - \alpha \, \frac{\partial J(w, \vec{b})}{\partial W_{11}^{(1)}} = 0.15 - 0.5(0.0004385) = 0.14978071$$

$$\frac{\partial J(w, \vec{b})}{\partial W_{12}^{(1)}} = i_2 \cdot \delta_1^{(2)} = 0.1 \times 0.008771 = 0.000877113$$

$$W_{12}^{(1)} := W_{12}^{(1)} - \alpha \, \frac{\partial J(w, \vec{b})}{\partial W_{12}^{(1)}} = 0.2 - 0.5 \, (0.000877113) = 0.19956143$$

$$\delta_2^{(2)} = \left(w_{12}^{(2)} \, \delta_1^{(3)} + W_{22}^{(2)} \delta_2^{(3)} \right) g' \left(z_2^{(2)} \right)$$

$$= \left[0.45 \times 0.1384 + 0.55 \times (-0.038098) \right] g' \left(0.25 \times 0.05 + 0.3 \times 0.1 + 0.35 \right) =$$

~~= 0.009954~~

$$\frac{\partial J(w, \vec{b})}{\partial W_{21}^{(1)}} = i_1 \cdot \delta_2^{(2)} = 0.05 \times 0.009954 = 0.000497712$$

$$W_{21}^{(1)} := W_{21}^{(1)} - \alpha \, \frac{\partial J(w, \vec{b})}{\partial W_{21}^{(1)}} = 0.25 - 0.5 \times 0.000497712 = 0.249751$$

$$\frac{\partial J(w, \vec{b})}{\partial W_{22}^{(1)}} = i_2 \cdot \delta_2^{(2)} = 0.10 \times 0.009954 = 0.0009954$$

$$w_{22}^{(1)} := W_{22}^{(1)} - \alpha \, \frac{\partial J(w, \vec{b})}{\partial W_{22}^{(1)}} = 0.3 - 0.5 \times 0.0009954 = 0.29950229$$